



(5') 1 SDLEQERRAKEKLQEQQ
18 SDLEQDRLAKEKLQEQQ
35 SDLEQERLAKEKLQEQQ
52 SDLEQERRAKEKLQEQQ
69 SDLEQERRAKEKLQEQQ
86 SDLEQDRLAKEKLQEQQ
103 SDLEQERRAKEKLQEQQ
120 SDLEQER~~LAKE~~KLQEQQ
137 SDLEQERLAKEKLQEQQ
154 SDLEQERRAKEKLQEQQ
171 SDLEQERRAKEKLQEQQ
188 SDLEQERRAKEKLQEQQ
205 RDLEQ

210 RKADTKKNLERKKEHGDILAEDLYGRLEIP
240 AIELPSENERGYYIPHQSISLPQDNRGNSRD
270 SKEISIIIEKTNRESITTNVEGRRDIHKGHL
300 EEKKDGSIKPEQKEDKS 316 (3') (SEQ ID NO:31)

FIGURE 1

(5') 1 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAAGTTGCAAGAACAAAC
52 AAAGCGATTAGAACAAAGATAGACTTGCCTAAAGAAAAGTTACAAGAGCAGC
103 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
154 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
205 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
256 AAAGCGATTAGAACAAAGATAGACTTGCCTAAAGAAAAGTTACAAGAGCAGC
307 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
358 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
409 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
460 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
511 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAAC
562 AAAGCGATTAGAACAAAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGC
613 AAAGAGATTAGAACAA
630 AGGAAGGCTGATACTGAAAAAAATTAGAAAGAAAAAGGAACATGGAGAT
681 ATATTAGCAGGAGTTTATGGTCGTTAGAAATACCAAGCTATAAGACTT
732 CCATCAGAAATGAACGTGGATTTATACCAACATCAATCTTCTTACCT
783 CAGGACAAACAGGGAAATAGTAGAGATTCCAAAGGAAATTCTATAATAGAA
834 AAAACAAATAGAATCTTACAACAAATGTTGAAGGGACGAAGGGATA
885 CATAAAAGGACATCTTGAAGAAAAGATGGTTCAATAAAACCAGAACAA
936 AAAGAAGATAATCT

950 (3') (SEQ ID NO: 32)

FIGURE 2

RDELFNELLNSVDVNGEVKENILEESQVNADIFNSLVKSVQQEQQ
HNVEEKVEESVEENDEESVEENVEENVEENDGSVASSVEESI
ASSVDESIDSSIEENVAPTVEEIVAPTVEEIVAPSVVEKCAPSVE
ESVAPSVEESVAEMLKER (SEQ ID NO: 24)

FIGURE 3

5' GAA TTC CGT GAT GAA CTT TTT AAT GAA TTA TTA AAT AGT GTA GAT
GTT AAT GGA GAA GTA AAA GAA AAT ATT TTG GAG GAA AGT CAA GTT AAT
GAG GAT ATT TTT AAT AGT TTA GTA AAA AGT GTT CAA CAA GAA CAA CAA
CAC AAT GTT GAA GAA AA AGT TGA AGA AAG TGT AGA AGA AA ATG ACG
AAG AAA GTG TAG AAG AAA ATG TAG AAG AAA ATG TAG AAG AAA ATG
ACG ACG GAA GTG TAG CCT CAA GTG TTG AAG AAA GTA TAG CTT CAA GTG
TTG ATG AAA GTA TAG ATT CAA GTA TTG AAG AAA ATG TAG CTC CAA CTG
TTG AAG AAA TCG TAG CTC CAA CTG TTG AAG AAA TTG TAG CTC CAA GTG
TTG TAG AAA AGT GTG CTC CAA GTG TTG AAG AAA GTG TAG CTC CAA GTG
TTG AAG AAA GTG TAG CTG AAA TGT TGA AGG AAA GGA ATT C 3' (SEQ ID No: 33)

FIGURE 4

FIGURE 5

NSRDSKEISIIEKTNRESITTNVEGRRDIHK
LSA-TER (seq ID NO: 23)
729S-NRI (seq ID NO: 26)
729S-NRII (seq ID NO: 27)
729S-Rep (seq ID NO: 28)

DEFNELLNVDUNGEVKENILEESQ
LEESQVNDDIFSNSLVKSVOQQEQQHNV
VEKCAPSVEESVAPSVEESVAEMLKER

NUCLEOTIDE SEQUENCE OF THE LSA GENE
5' END

(NON-CODING 5' END)

1 AAAGTATACATCTTCCTTCTTACTTCTTAAA

(CODING 5' END, UNIQUE)

33 ATGAAACATATTTGTACATATCATTACTTATCCTGTTAATTATTG
84 ATATTCAATAATGAAAGATAATAAGAATTCTGAAAAAGATGAAATCA
135 TAAAATCTAACTTGAGAAGTGGTCTTCAAATTCTAGGAATCGAATAATGA
186 GGAAAATCACGAGAAGAACACGTTTATCTCATAATTATGAGAAAAC
237 AAAAATAATGAAATAATAAATTTCGATAAGGATAAAAGAGTTAACGATGT
288 CTAATGTAAGGAAATTGTGTACAAACAATTCAAAAGTCTTTAAGAAATCT
339 TGGTGTTCAGAGAATATATTCTTAAAGAAAATAATTAAAGGAAGGG
390 AAATTAAATTGAACACATAATAATGATGATGACGATAAAAAAAATATTA
441 AAGGGCAAGACGAAAACAGACAAGAAGATCTGAAGAAAAGCA

(CODING 5' END, repetitive)

492 GCTAAAGAAAAGTTACAGGGCAACAAAGCGATTCAGACAAGAGAGACGT
543 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACCT
594 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACGT
645 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACCT
696 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACGT
747 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACGT
798 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACCT
849 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGACAAGAATAGACCT
900 GCTAAAGAAAAGTTGCAAGAACAAACAAAGCGATTTAGACAAGAGAGACGT
951 GCTAAAGAAAAGGTTGCAAGAACAAACAAAGCGATTTAGA 988 (SEQ ID NO: 34)

FIGURE 6

LSA. 5' /ATG - -> 1-phase Translation

(Seq ID No: 35)

DNA sequence 956 b.p. ATGAAACATATT ... AAGCGATTAGA linear

(Seq ID No: 36) 1 / 1 31 / 11
ATG AAA CAT ATT TTG TAC ATA TCA TTT TAC TTT ATC CTT GTT AAT TTA TTG ATA TTT CAT
(Seq ID No: 36) met lys his ile leu tyr ile ser phe tyr phe ile leu val asn leu ile phe his
61 / 21 91 / 31
ATA AAT GGA AAG ATA AAG AAT TCT GAA AAA GAT GAA ATC ATA AAA TCT AAC TTG AGA
ile asn gly lys ile ile lys asn ser glu lys asp glu ile ile lys ser asn leu arg
121 / 41 151 / 51
AGT GGT TCT TCA AAT TCT AGG AAT CGA ATA AAT GAG GAA AAT CAC GAG AAG AAA CAC GTT
ser gly ser ser asn ser arg asn arg ile asn glu glu asn his glu lys lys his val
181 / 61 211 / 71
TTA TCT CAT AAT TCA TAT GAG AAA ACT AAA AAT GAA AAT AAA TTT TTC GAT AAG
leu ser his asn ser tyr glu lys thr lys asn asn glu asn lys phe phe asp lys
241 / 81 271 / 91
GAT AAA GAG TTA ACG ATG TCT AAT GTA AAA AAT GTG TCA CAA ACA AAT TTC AAA AGT CTT
asp lys glu leu thr met ser asn val lys asn val ser gln thr asn phe lys ser leu
301 / 101 331 / 111
TTA AGA AAT CTT GGT TCA GAG AAT ATA TTC CTT AAA GAA AAT AAA TTA AAT AAG GAA
leu arg asn leu gly val ser glu asn ile phe leu lys glu asn lys leu asn lys glu

FIGURE 7A

NUCLEOTIDE SEQUENCE OF THE LSA GENE
3' END

(CODING 3' END, REPETITIVE)

1 CAAGAACAAACAAAGCGATCTAGAACAAAGAGAGACGT
37 GCTAAAGAAAAGTTCAAGAACAACAAAGCGATTTAGAACAGATAGACTT
88 GCTAAAGAAAAGTTCAAGAACAACAAAGCGATTTAGAACAGAGAGACTT
139 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATCTAGAACAGAGAGACGT
190 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGAGAGACGT
241 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGATAGACTT
292 GCTAAAGAAAAGTTCAAGAAGCAGCAAAGCGATTTAGAACAGAGAGACGT
343 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGAGAGACGT
394 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGAGAGACTT
445 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGAGAGACGT
496 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGAGAGACGT
547 GCTAAAGAAAAGTTCAAGAACAAACAAAGCGATTTAGAACAGAGAGACGT
598 GCTAAAGAAAAGTTCAAGAGCAGCAAAGAGATTTAGAACAA

(CODING 3' END, UNIQUE)

640 AGGAAGGCTGATACGAAAAAAAATTAGAACAGAAAAGGAACATGGAGAT
691 ATATTAGCAGAGGATTTATGGTCGTTAGAAATACCAGCTATAGAACCT
742 CCATCAGAAAATGAACGTGGATATTATACCATCAATCTTCTTACCT
793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAA
844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATA
895 CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA
946 AAAGAAGATAAATCTGCTGACATACAAATCATACATTAGAGACAGTAAT
997 ATTTCTGATGTTAATGATTTCAAATAAGTAAGTATGAGGATGAAATAAGT
1048 GCTGAATATGACGATTCTTAATAGATGAAGAACAGATGATGAAGACT
1099 TAGACGAATTAAAGCCTATTGTGCAATATGACAATTCCAAAGATGAAGAAA
1150 ACATAGGAATTATAAAGAACTAGAACGATTGATAGAGAAAATGAAAATT
1201 TAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAACAGATTATCTGAAG
1252 AAAAAATAAAAAAGGAAAGAAATATGAAAAACAAAGGATAATAATTTA
1303 AACCAAATGATAAAAGTTGTATGATGACATATTAAAAATATAAAATG
1354 ATAAGCAGGTTATAAGGAAAAGGAAAATTCAAAATCATTGTTTCATA
1405 TATTTGACGGAGACAATGAAATTTCAGATCGTGGATGAGTTATCTGAAG
1456 ATATAACTAAATATTTATGAAACTATAA (stop) (SEQ ID NO: 39)

(NON-CODING 3' END)

1485 AAGGTTATATATT 1498

FIGURE 8

LSA. 3'. ALL -> 1-phase Translation
 (Seq ID No: 40) (Seq ID No: 41)
 DNA sequence 1496 b.p. CAAGAACAAACAA .. GGTTATATATT linear

1 / 1	31 / 11
(Seq ID No: 42) CAA GAA CAA CAA AGC GAT CTA GAA CAA GAG AGA CGT GCT AAA GAA AAG TTG CAA GAA CAA	
{Seq ID No: 43} gln glu gln gln ser app leu glu gln glu arg arg ala lys glu lys leu gln glu gln	
61 / 21	91 / 31
CAA AGC GAT TTA GAA CAA GAT AGA CTT GCT AAA GAA AAG TTA CAA GAG CAG CAA AGC GAT	
gln ser app leu glu gln asp arg leu ala lys glu gln glu gln ser app asp	
121 / 41	151 / 51
TTA GAA CAA GAG AGA CTT GCT AAA GAA AAG TTG CAA GAA CAA AGC GAT CTA GAA CAA	
leu glu gln glu arg leu ala lys glu leu gln glu gln ser app asp leu gln	
181 / 61	211 / 71
GAG AGA CGT GCT AAA GAA AAG TTG CAA GAA CAA AGC GAT TTA GAA CAA GAG AGA CGT	
glu arg arg ala lys glu lys leu gln glu gln gln ser app asp leu glu gln glu arg arg	
241 / 81	271 / 91
GCT AAA GAA AAG TTG CAA GAA CAA CAA AGC GAT TTA GAA CAA GAT AGA CTT GCT AAA GAA	
ala lys glu lys leu gln glu gln gln ser app asp leu glu gln asp arg leu ala lys glu	
301 / 101	331 / 111
AAG TTA CAA GAG CAG CAA AGC GAT TTA GAA CAA GAG AGA CGT GCT AAA GAA AAG TTG CAA	
lys leu gln glu gln ser app asp leu glu gln glu arg ala lys glu lys leu gln	

FIGURE 9A

1261 / 421	1291 / 431
AAA GGA AAG AAA TAT GAA ACA ACG GAT AAT TTT AAA CCA AAT GAT AAA AGT TTG	
lys gly lys tyr glu lys thr lys asp asn pro asp lys ser leu	
1321 / 441	1351 / 451
TAT GAT GAG CAT ATT AAA AAT GAT AAG CAG GTT AAT AAG GAA AAG GAA AAA	
tyr asp glu his ile lys lys tyr lys asn asp lys gln val asn lys glu lys	
1381 / 461	1411 / 471
TTC ATA AAA TCA TTG TTT CAT ATA TTT GAC GGA GAC AAT GAA ATT TTA CAG ATC GTG GAT	
phe ile lys ser leu phe his ile phe asp gly asp asn glu ile leu gln ile val asp	
1441 / 481	1471 / 491
GAG TTA TCT GAA GAT ATA ACT AAA TAT TTT ATG AAA CTA TAA AAG GTT ATA TAT	
glu leu ser glu asp ile thr lys tyr phe met lys leu [och lys tyr tyt]	

* Strike from Fig.

FIGURE 9D

LSN. 3' STOP -> 1-phase Translation (SEQ ID No: 44) (SEQ ID No: 45)
 DNA sequence 1482 b.p. CAAGAACAAACAA ... ATGAAACTATAA linear

FIGURE 10A

1321 / 441 1351 / 451
TAT GAT GAG CAT ATT AAA TAT AAA AAT GAT AAG CAG GTT AAT AAG GAA AAG GAA AAA
tyr asp glu his ile lys lys tyr lys asp lys asn lys glu lys glu lys
1381 / 461 1411 / 471
TTC ATA AAA TCA TTG TTT CAT ATA TTT GAC GGA GAC AAT GAA ATT TTA CAG ATC GTG GAT
phe ile lys ser leu phe his ile phe asp asn glu ile leu gln ile val asp
1441 / 481 1471 / 491
GAG TTA TCT GAA GAT ATA ACT AAA TAT TTT ATG AAA CTA TAA AAG GTT ATA TAT
glu leu ser glu asp ile thr lys tyr phe met lys leu ~~tyr~~
*Delete from figure

FIGURE 10D